Precision Water Management in Corn Using Automated Crop Yield Modeling and Remotely Sensed Data

Sudhanshu Sekhar Panda[1], Dean D. Steele[2], Suranjan Panigrahi[2], Daniel P. Ames[3]

[1]Department of Science, Engineering, and Technology Gainesville State College, 3820 Mundy Mill Road, Idaho, USA

[2]Agricultural and Biosystems Engineering Department North Dakota State University, Idaho, USA

[3]Department of Geosciences, Idaho State University, Idaho, USA

spanda@gsc.edu

***Abstract*-A model of crop yield versus seasonal water use was developed based on late July and early August vegetation aerial image data; bare soil image data; land elevation data; climatic data (temperature, accumulated growing degree days, solar radiation, rainfall); and non-climatic non-imagery data (irrigation application). All parameters were integrated from the germination date to the aerial image acquisition date and a radial basis functional network yield prediction model was developed. The resulting model provided an average prediction accuracy of 91% with a correlation coefficient (*r*) of 0.65. The standard error of prediction (SEP) and Root Mean Square Error (RMSE) obtained from the model was only 9.62% and 10.2% of the average actual yield of the test dataset. A linear fit model was created using the spatially predicted corn yields versus the corresponding estimated ET for the crop. An R^2 of 0.65 was obtained from the model. A studentized residual test and Q-test suggested several probable outliers in the test data. After the elimination of these outliers, the linear fit model between estimated ET and predicted corn yield provided an improved R^2 of 0.81. It is expected that farmers and analysts could use the developed water use model to estimate the seasonal water requirement for corn in a midseason cropping period.**

Keywords-Evapotranspiration, radial basis function network (RBFN), Aerial imaging, Principal component analysis (PCA), Water use model, Outlier analysis

I. INTRODUCTION

Efficient crop production and high yields depend upon the best use of available water (Derenbos and Pruitt, 1977) A method is required to measure the seasonal water requirement for a crop according to a yield goal. According to Klocke et al. (1996), irrigators need to learn the means to convert water into grain in the most efficient manner possible. Applying only enough water to meet full evapotranspiration (ET) of the crop (or crop water use) is one of the keys to efficient water use. Since ET is directly related to crop yield, efficient water management can be achieved by supplementing rainfall with enough irrigation water to meet the full water requirement. Irrigation management influences production costs and affects leaching of nutrients to groundwater (Steele et al., 2000). However, while water stress reduces crop yield (Ali et al., 2007), over-irrigation causes runoff or percolation of excess water beyond the root zone, the latter which can leach nitrogen and pesticides into ground water. Runoff can also be polluted in this process.

Irrigation is becoming more expensive as energy costs increase, aquifers become depleted, and irrigation water quality decreases (Terjung et al., 1984). Typically, irrigation applications greatly exceed crop water requirements, if not carefully planned.,Use of a Plant Growth – Water balance model can be an effective approach to such planning (Gallardo et al., 1997; Stegman, 1986). Water demand from the non-agricultural sector along with adverse weather cycles are forcing the agriculture industry to devise new technologies to improve water management capabilities and promote efficient use of water resources (Xin et al., 1997). Water stress has economic implications because it produces low crop yield (Steele et al., 2000). Hence, it is essential to estimate the crop yield before harvest to improve the management of water resources. Therefore, yield versus water use models are necessary for different crops to increase productivity while using water resources efficiently.

Several studies have reported (reviewed by Hanks, 1983) on linear corn yield responses (dry matter or grain) related to water use. Stegman (1982), Derenbos and Pruitt (1977) and Klocke et al. (1996) found a linear relationship between the corn grain yield and cumulative ET for the crop-growing season. Stegman (1982) reported that maximum grain yield was consistently produced with low water stress sequences in the entire crop season.

Soil properties and crop yields are often spatially correlated (Burrough et al., 1993; Timlin et al., 1998). Soil texture can be determined from bare soil remotely sensed images with acceptable accuracy (Lillesand and Kiefer, 1995). Topography can also be a major factor that affects the crop yield on a spatial basis. When surface soil depth decreases, the fertility of the land reduces with organic matter loss, which in turn affects the crop yield. Water stress is also a cause of low crop yield (Steele et al., 2000). Therefore, the use of all these crop production parameters, i.e., early stage crop vigor reports from remotely sensed vegetation images, soil texture information from bare soil images, topography, seasonal nitrogen and irrigation applications, and crop season weather information including rainfall, may be useful in the prediction of crop yield. Terjung et al. (1984) observed that various climatic and non-imagery crop production features have been used along with image information to predict crop yield.

Zhuang and Engel (1990) and Ranaweera et al. (1995) provided research evidence demonstrating the superiority of artificial neural networks (ANN) over statistical models in nonlinear data modeling, diminishing collinearity problems, and increasing model flexibility. Considering the non-linear relationships of crop yield with the affecting parameters, ANNs based on artificial intelligence may be a better substitute. ANNs have the ability of computing, processing, predicting and classifying data, and have the advantage of nonlinearity, input-output mapping, adaptivity, generalization, and fault tolerance (Haykin, 1999).

Radial basis function neural networks (RBFN) have found increased application across many disciplines (Bishop, 1995) due to their structural simplicity and training efficiency (Lee et al., 1996; Goodman, 1993; Wan and Harrington, 1999). RBFNs have a guaranteed learning algorithm, e.g., linear least squares optimization. There is an extra prototype layer in the RBFN architecture, where the input vectors (after being multiplied with random weights) are clustered into different homogeneous groups. The clustered centroids are fed to the prediction model for output prediction (Haykin, 1999) and thus make it a better prediction NN model. Therefore, despite its computational intensity and its optimal parameterization problem (Wan and Harrington, 1999), RBFN is recently becoming popular and may also be used for smaller datasets such as that used for this study.

The objectives of this study were twofold: i) to predict corn yield on a spatial basis using crop management information, environmental parameters, and RBFN modeling techniques, and ii) to develop a linear seasonal water requirement model based on the spatial basis predicted yield. This study predicts the corn yield spatially in midseason (more than 10 weeks before harvest) and, subsequently, the seasonal water requirement from the predicted yield versus water use model. Farmers may benefit from estimates of seasonal water requirements made (midseason) based on forecasted yield, so that they can take corrective measures during the remainder of the season for either enhancing crop growth or reducing excessive irrigation. Irrigation scheduling can be done in accordance with the rainfall for the remainder of the growing season. Very few studies have been completed in predicting seasonal water requirements for crops. Prediction of seasonal water requirements could reduce the amount of excess water applied. Simultaneously, it can help the farmers to reduce the spatially varying water stress in the field, so that crop yield can be enhanced to an optimum level.

II. MATERIALS AND METHODS

A. *Study Area Description*

Farmers are accustomed to field scale agriculture. According to Steele et al. (2000), with the inclusion of commercial production constraints such as time, labor, and energy, field-scale research has advantages over plot (controlled)-scale research. Field scale research also integrates across field-scale soil variability (Steele et al., 2000), and may appeal to farmers as a tool due to its production-sized nature. Field forecasted corn yield is useful input information to estimate seasonal water requirement. Therefore, irrigation scheduling could be made to reduce wastage of water and help detect water stress areas, suggesting corrective measures to enhance yield.

Our study area is the best management practice (BMP) site in the Oakes Irrigation Test Area (OITA) near Oakes, North Dakota, USA. The BMP site is a 65 ha field located at 46.051974 0 N, -98.111879 0 E, and 399 m elevation above mean sea level. The study was conducted using four years' of field data, i.e., 1997, 1998, 2000, and 2001. A sub-humid climate prevails at the Oakes BMP site with an average precipitation of 310 mm for May through September. Freeze-free periods average 135 days with accumulation of growing degree-days (GDD) (base 10 0C) of 1182 (Stegman, 1982). The average rainfall and GDD amount from May to September for 1997 to 2001 are provided in Table 1. Meteorological data were measured at an automated weather station within 2.4 km of the site. Rainfall was measured at the site.

The soil in the south half of the field is predominantly Hecla loamy fine sand (sandy, mixed, frigid OxyaquicHapludoll); in the north Wyndmere fine sandy loam (coarse-loamy, mixed, superactive, frigid AericCaciaquall) and Stirum fine sandy loam (coarse-loamy, mixed, superactive, frigid TypicNatraquolls) dominate (Derby et al., 1998). The field was irrigated with a center pivot system with end guns. The irrigation system capacity is approximately 59 l s-1 and can apply approximately 30 mm of irrigation water in 72 h for a 3600 revolution (Steele et al., 2000). The quarter section is divided into four quadrants for irrigation scheduling purposes and designated northwest, northeast, southwest, and southeast (NW, NE, SW, and SE, respectively). The four corners of the field within a distance of 200 meters from the corner point (both sides) were not irrigated (Figure 1). Corn was grown in the SW and SE quadrants in 1997, in the entire field in 1998, in the SW and SE quadrants in 2000, and in the NW and NE quadrants in 2001. The corn variety in all these years was always Pioneer 3751. The row spacing was 60 cm. The planting density in all years was 74,100 per ha. In all four years, pre-plant applications of N, P, and K were made each season to eliminate these nutrients as production limiting factors. Nitrogen application was carried out in three different stages of the crop season. They were pre-plant, side dress, and fertigation stages. The agronomics of the four crop seasons is shown in Table 1.

B. *Aerial Image Acquisition and Image Processing*

False color composite (FCC) aerial images of mid-season cropping periods in each year were acquired from the BMP site. In this study, the images were taken using a broad range of visible spectrum, ranging from 400 nm to 700 nm. Visible band (R, G, and B) aerial images (Table 1) were used in the study. Figure 1 shows a typical vegetation image of the OITA study area. Color calibration of the aerial images for different years was performed using image pixels from colored (Red, Green, Blue, White, and Black) sheets as reference. The camera used for the image acquisition was calibrated to an ideal standard in the laboratory before being used to reduce aberration in image gray levels. All the images were also acquired in cloud free conditions around noon (for similar sun angles). Other image acquisition parameters, such as flight height and image acquisition systems, were kept the same for all the image acquisition dates. The images in all four years

(1997, 1998, 1999, and 2001) were acquired during a narrow window of growing season or at similar growth periods in different years. Therefore, it was assumed that the effect of radiometric aberrations in the aerial images were minimum.

The aerial imaging system used for the image acquisition was a SLM 35mm camera loaded with either 100 or 200 ASA Ektachome slide film. The film was developed into photographs and later scanned with a Nikon Scanner at 2800-dpi (dot per inch) resolution. The aerial images were saved in 8-bit TIFF format. The resolution was 0.6 m x 0.6 m. Raw images were not geometrically corrected. The images were georeferenced with geographical ground coordinates of each corner of the rectangular plot (already recorded ground control points and shown in Figure 1). Although soil information is a non-imagery factor affecting crop yield, aerial images of bare soil were used to represent the soil factors responsible for crop yield variation. The dates of bare soil images used in this study are provided in Table 1.

C. *Spatial Management for Input Data Acquisition*

Twenty undisturbed lysimeters were installed in the site prior to the 1997 growing season. Among them, four lysimeters each were installed in each of the four quadrants. An additional four lysimeters were installed in each of the non-irrigated corners (Figure 1). Precipitation was recorded at each lysimeter. Smaller grid images of 65 x 80 pixels were extracted around individual lysimeters. Figure 1 represents the lysimeters' position along with their numbers. The grid plots are also marked in the same figure. Grid images were extracted from the corresponding bare soil images relating to the spatial coordinates.

In 1997, corn was grown in the SW, SE and the non-irrigated corners of the field. Twenty (8 + 8 + 4) grid images were extracted from the aerial image. The maximum possible 36 grid images were extracted from the study area in 1998. Twenty grid images were extracted from the SW and SE quadrants of the study area in 2000. Only 16 grid images were extracted from NW and NE quadrants' aerial images in 2001. Thus, a total of 92 images were available for the development of the yield prediction model.

Non-imagery data, such as elevation, was measured using a surveying transit on a rectangular grid; the measurement interval was 20 m in each direction. The elevation was interpolated using the kriging geostatistics procedure with Surfer 32 (Golden Software, Golden, CO), and the average elevation of each grid plot was determined. The meteorological data represented the entire site in a particular year. Averages of the maximum and minimum air temperatures (from the planting date to the mid-season image acquisition date), corn accumulated growing degree-days (AGDD) (from the date of planting to the image acquisition date), and available water (irrigation and precipitation) to the plant (from the plant emergence to the image acquisition date) were also collected as the non-imagery environmental parameters. Average solar radiation for the same periods was also collected for all four years. The crop emergence dates, the plant maturity dates, aerial image acquisition dates, and the last non-freezing dates of the crop seasons are provided in Table 1. The use of crop production features from the planting date to the aerial image acquisition date would help framers predict crop yield for any ensuing year.

TABLE 1. AGRONOMIC SUMMARY

Agronomic Items	**1997**	**1998**	**2000**	**2001**
Spring Soil Test N to 0.6 m (kg/ha)	28	54	37	34
Preplant N (kg/ha)	10 south	24	11	10 starter + 112 (variable rate)
Planting date	10May south	25-Apr	24-Apr	8-May
Corn Variety	Pioneer 3751 south	Pioneer 3751	Pioneer 3751	Pioneer 3751
Plant Population (ha^{-1})	74,100	74,100	74,100	74,100
Emergence	27-May	10-May	6-May	15-May
Sidedress Date	23-June	26-June	26-June	n/a
Sidedress amount (kg N ha^{-1})	168	134	157	n/a
Fertigation amount (kg N ha^{-1})	56	45	84	56
Plant maturity date	1 October	9 September	15 September	26 September
Image acquisition date (vegetation)	2 August	30 July	29 July	2 August
Last non-freezing date	13 October	1 October	15 September	4 October
Image acquisition date (soil)	15 May	28 April	16 May	26 May
Harvest date	28 October	23 October	13 October	23 October

D. Spatial Yield Data Collection

The grid based yield was recorded along with longitude and latitude using the "Micro Track" program (Micro Track Software Corporation, Wyomissing, PA) with a yield monitorof 6 m intervals. Data were transformed to a new matrix form of 3 m intervals by kriging. We were able to fill in the missing data in the dataset and extract yield information for the grid images (65 x 80 pixels). We extracted the average yield from each plot (grid image) of size 39 m x 48 m using the average sampling algorithm, expressed as

$$Y_{GP} = \frac{\sum_{i=1}^{n} X_i}{n}, \quad (1)$$

where YGP is the average crop yield from an individual plot (corresponding to grid images), Xi is the yield from each individual 3 m grid within the 39 m x 48 m plot, and n is the total number of individual 3 m grids present in the entire plot. Spatial coordinates of the grid-based yield were used to correlate with the grid images. Actual yield from the field was used as the output parameter (neuron or processing element) for the RBFN model. Figure 2 represents the input parameters and the output parameter for the neural network model.

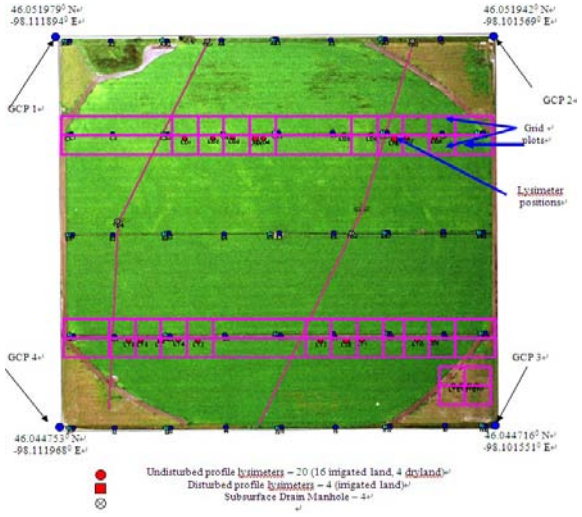

Figure 1. A typical aerial (vegetation) image (August 2, 1997) of the field site with lysimeyters and grid plot positions.

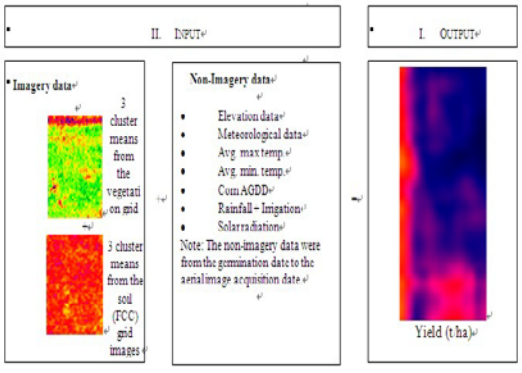

Figure 2. The input features and output feature used for the neural network yield prediction model building.

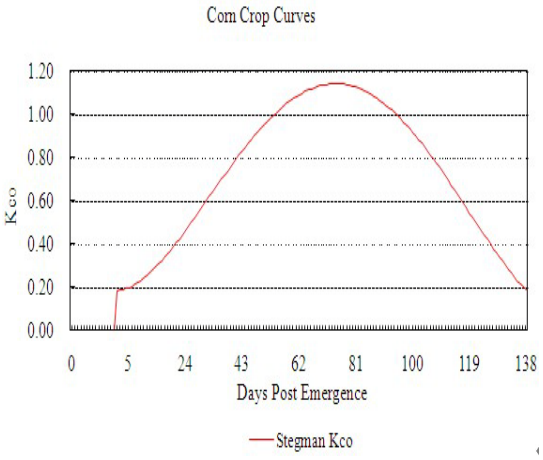

Figure 3. The basal crop coefficient curve for calculation of corn ET (from Stegman and Coe, 1984)

E. Image Processing

Data from all spectral bands include a certain degree of redundancy (Fung and LeDrew, 1987). Principal component analysis (PCA) analyzes correlated multidimensional data (Byne et al., 1980). PCA is a linear transformation that reorganizes the variance in a multi-band image into a new set of image bands. Each individual band in the output PCA image receives some contribution from all of the input image's bands. Therefore, the principal component band accounting for highest variability (more than 82%) of the individual band's gray scale information was used for this study. Principal component band images for both vegetation and soil images for all four years were extracted. For all the multi band images the first principal component band (PC1) contained more than 82% information.

To overcome the inconvenience in choosing training sites for a large number of images and to make the classification process simpler and less time consuming, a technique of clustering homogeneous spectral information (unsupervised classification) was used rather than a supervised classification technique. In this case, no advance information about the classes of interest was required. The iterative self-organizing data analysis (ISODATA) clustering technique was used to cluster the PC1 images. Based on overall observations, we decided to use three clusters for vegetation images. Bare soil images were also clustered into three distinct clusters that corresponded to the three soil classes at the site. In some of the grid plots, all three soil classes were not present. ISODATA clustering technique showed the absence of these clusters with either very few pixels or no pixels in the corresponding groups.

F. ET Data Collection Method

Seasonal crop water usage was calculated as the summation of daily crop water use, which was estimated by using the Jensen and Haise equation (1963) and a set of modifying coefficients (Stegman and Coe, 1984). Potential ET was based on weather data, specifically, maximum and minimum daily air temperatures (0C) and solar radiation (MJ m-2). The above-mentioned climatologic parameters were the same for all the quadrants and the non-irrigated corners for each year. However, the irrigation amount varied in each quadrant for each year.

The individual lysimeters in the quadrants recorded the variation in irrigation amount on a spatial basis. Potential ET from each quadrant (the selected grid-plots around each lysimeter) as well as from the non-irrigated corner was calculated for each year. This was calculated using the following Jensen-Haise (1963) equation:

$$ET_r = 0.0102(T_m + 3.36)\, R_s \quad (2)$$

whereETr = reference evapo-transpiration (mm d-1), Tm = average daily temperature (in 0C) and calculated as Tm = (Tmin +Tmax)/2, and Rs = daily solar radiation (MJ m-2 d-1).

Daily crop evapo-transpiration (ET) (mm) was estimated as:

$$ET = K_c ET_p, \quad (3)$$

whereKc is the empirically determined crop coefficient, when water supply fully meets the water requirement of the crop. The value of Kc varies with crop and developmental stage. Kc takes a different form for the Jensen et al. (1971) scheduling model. This can be expressed as

$$K_c = K_{co} K_a + K_s \quad (4)$$

whereK_{co} = coefficient used for potential ET modification in plant growth stage, K_a = coefficient (ranging from 0 to 1) of the soil moisture deficit (SMD), and K_s = coefficient to increase crop ET when the soil surface is wet after rainfall and /or irrigation (Stegman and Coe, 1984). Figure 3 represents a typical K_{co} curve based on Stegman and Coe (1984).

For this study, Kco was based upon curves developed for corn in North Dakota (Stegman et al., 1977). These curves are fitted by fourth order polynomials, i.e.,

$$K_{co} = C_1 + C_2\, DPE + C_3\, DPE^2 + C_4\, DPE^3 + C_5\, DPE^4 \quad (5)$$

where DPE is days past emergence and the coefficients are as follows: C_1= - 0.1814466119, C_2= 1.877271 x 10^{-4}, C_3= 7.004694 x 10^{-4}, C_4= - 9.3707 x 10^{-6}, and C_5= 3.12 x 10^{-8}.

A mathematical artifact of equation (5) is that the curve increases after 140 days past emergence, implying increasing crop water use late in the season. To avoid this difficulty, simulations were not extended beyond the date of crop maturity.

Stegman et al. (1977) reported the formula to determine the K_a factor that was used for the correction of limiting soil moisture condition. The equation used is given by:

$$K_a = 1, \quad if \quad AW > 50\% \quad (6a)$$

$$K_a = AW / 50, \; if \; AW < 50\% \quad (6b)$$

where AW = percent available water remaining (100 when root zone is at field capacity). The adjustment for wet surface soil conditions was limited to 3 days after rainfall or irrigation in periods when green ground cover was incomplete, i.e., K_{co}< 0.9 (Stegman et al., 1977). Therefore, we used different Ks factors based on different soil moisture conditions due to either rainfall or irrigation.

There were no soil moisture deficits (SMD) exceeding 50% in 1997, while in 1998 there were 14 consecutive days in which SMD ranged from 50 to 65%. The SMD in 2000 was for a period of 54 days, which ranged from 50 to 95%. In 2001, 19 days were observed in which SMD ranged from 50 to 65%.

G. ET Calculation

Rain gauges were placed in each quadrant to measure rainfall and irrigation amounts, including the non-irrigated corner areas to record rainfall. Rainfall and irrigation amounts were separated based on the rainfall reading acquired by rain gauges in the non-irrigated corners. Temperature, solar radiation, AGDD, and reference evapotranspiration (ET_0) were considered uniform for the entire field site and obtained from the North Dakota Agriculture Weather Network (NDAWN) weather station located 2.4 km from the site. Each rainfall event was treated as constant for each quadrant in the study area, while irrigation amount, soil moisture level, drainage amount, ET_0, and crop yield were constants for each lysimeter (i.e., each grid image) position. We used the climatologic and non-climatologic information of the non-freeze period of crop season for each year (Table 1) and rainfall and irrigation amount from plant emergence to the plant maturity dates (Table 1).

Evapotranspiration values for each lysimeter were calculated using the algorithms described previously. We obtained 20 different ET values for corresponding grid images for each year. Only the ET values of the lysimeters where corn was grown were considered for each year.

H. Neural Network Model Building

According to Moody and Darken (1989), a typical RBFN consists of three different layers, with each successive layer fully connected by feed forward arcs as shown in Figure 4. There is no provision of weight between input layer and the hidden layer (prototype). The transfer function used at the hidden layer is the "radial basis function," which is a nonlinear transfer function. There is only one hidden layer present, which is fully connected to a linear output layer (Neural Ware, 2000).

The radial basis function is defined by the following equation.

$$F(x) = \sum_{i=1}^{N} w_i \varphi(\| x - c_i \|), \quad (7)$$

where$\{\varphi(||x - xi||)\}\; i = 1, 2, \ldots, N\}$ is a set of N arbitrary functions, x is the input data vector, w_i is the randomized weight vector, ‖ . ‖ is an Euclidean distance measure, and the ci is the known data point considered to be the center of the radial basis functions.

Therefore, the output of the i^{th}-hidden neuron can be written as

$$h_i(x) = -0.5 * \frac{(\| x - c_i \|)^2}{\delta_i^2}, \quad (8)$$

where$h_i(x)$ is the pattern units in the hidden layer that is connected to the output layer, $\boldsymbol{x}$ is the input vector, c_i is the cluster center of the radial basis function, and δ is the common Gaussian function. In the Neural Ware Professional II Plus, the learning of the pattern units consists of using a clustering

algorithm known as k-means clustering and the nearest neighbor heuristic to determine δ_i.

As the output layer is linear, the total output is the summation of outputs of all hidden neurons connected to the output neuron. Thus, for the j^{th} neuron input layer, the output y_j corresponding to the input vector x is

$$y_j(x) = \sum_{i=1}^{k} W_{ji} h_i(x) + b_k, \qquad (9)$$

where W_{ji} is the synaptic weight connecting hidden neuron i to output neuron j, and k is the number of hidden layer neurons. The bias b_k is added to the output. No extra hidden layer is used in our study other than the hidden layer with the radial basis function.

For this study, Moody and Draken RBFN (MD-RBFN) transfer function (Moody and Draken, 1989), standard k-means clustering, and nearest neighborhood heuristic were used to determine the cluster center ci and Gaussian weights δ_i(Neural Ware, 2000).

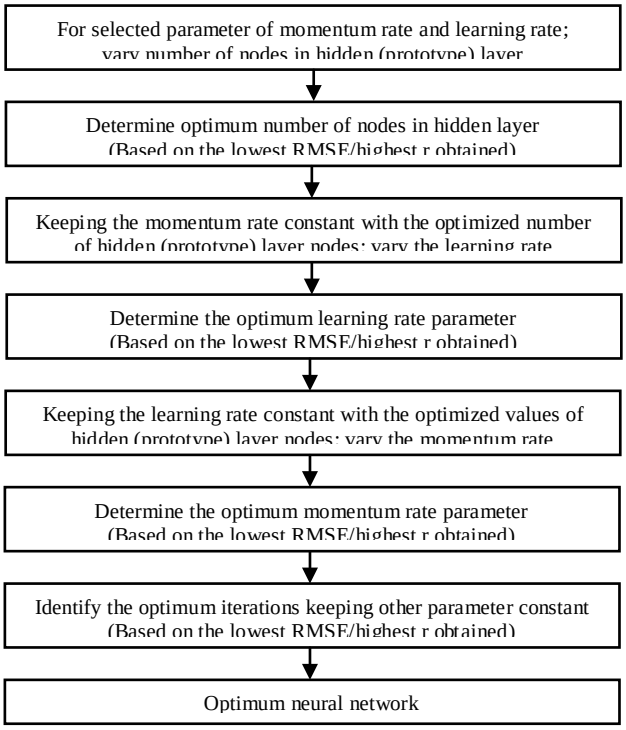

Figure 4.Structure of a typical radial basis function network.

We obtained three cluster average values from ISODATA clustering of PC1 band-vegetation grid images, three cluster averages from the PC1 band-soil images, average elevation data, and other climatic data (Figure 3) to use as input neurons to the RBFN in the Neural Ware Professional Plus II software (NeuralWare, Carnegie, PA). Corresponding average yields from each grid plot were the output neurons. Thus, the RBFN model had 11 input neurons and one output neuron (corn actual yield). The model used all four years (1997, 1998, 2000, and 2001) combined data to select the training and testing data on a random basis. We used 60 training observations and 32 testing observations. The test data set were chosen to represent each lysimeter (for which the corn crop yields were available) covering all four years (Figure 4).

Data transformation techniques are useful to enhance the input and output correlation. These techniques reduce the range difference in data points and can make the entire input dataset uniform. We used log transformation ($\log_{10}$X) to transform the high range inputs such as elevation and corn AGDD.

The initial RBFN network was defined with learning rate of 0.5, momentum term of 0.4, ten prototype layers, 20000 epochs, delta learning rule, and sigmoid transfer function. A step-by-step optimization procedure (Figure 5) was followed to optimize the parameters of the models. TanH transfer function was also used to compare the model performance. The RBFN model performances were evaluated based on Root Mean Square Error (RMSE), prediction accuracy, and standard error of prediction (SEP).

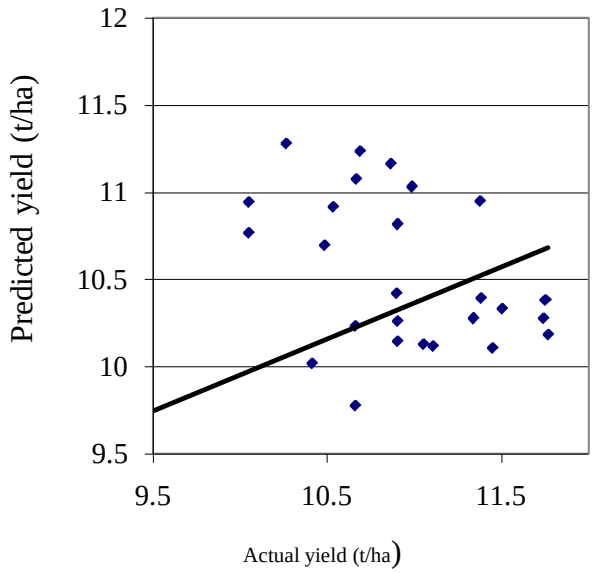

Figure 5.Schematic of procedure for determining the optimum neural network architecture.

The equation for RMSE is given by

$$\text{RMSE} = \sqrt{MSE} = \sqrt{\frac{SSE}{n}}, \qquad (10)$$

where n is number of observations, *SSE* and *MSE* are sum of squared error and mean square error, respectively. Average test prediction accuracy is calculated based on the equation

$$\textit{Average Test Accuracy}\,(\%) = \left(1 - \frac{1}{n}\sum_{i=1}^{n}\frac{|Y - X|}{Y}\right) x100, \qquad (11)$$

where n is total number of observations and Y and X are actual and predicted output, respectively.

The SEP of the predictive model is calculated by the equation (Kramer, 1998).

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}[(Y_i - X_i) - d_m]^2}{n-1}}, \quad (12)$$

where d_m is mean of the difference between actual and predicted values Y and X (of i^{th} individual), respectively, and n is the total number of observations.

I. *Crop Yield Versus Water Use Modeling and Output Residual Analysis*

Although grid-plots surrounding the lysimeters in each quadrant had equal ET, their corresponding yields were different because the type of soil surrounding the lysimeters was considered uniform. Therefore, we used the same ET value within each quadrant for water use model building. We developed a linear regression model with predicted yield as the independent variable and calculated ET (from the ET calculation model) as the dependent variable.

To detect possible deficiencies in the prediction model, a residual analysis was performed with the model outputs using i) studentized t-test, ii) Q-test, and iii) confidence interval analysis (Canovas, 1984). The plot of the studentized concentration residuals versus the leverage value of a data set provides information on probable outliers. The data point is a potential outlier if the sample has both high leverage and a high studentized residual. Leverage value gives a measure of how important an individual in a data set used in model building. The leverages for the data points are calculated by the following equations.

$$H = X\,(X'X)^{-1}\,X' \quad (13)$$

$$Leverage_i = H_{ii} \quad (14)$$

where X is the data matrix (n by f) of sample values, H is the n by n square matrix, n is the number of samples in the data set, and f is the number of factors in the model. In this case, n is 32 as we had 32 estimated ET relating to the grid plots used as the testing dataset for corn yield prediction. The values H_{ii} are the diagonal elements in the square matrix H. The subscript i is the sample number in the data set. The studentized residual is calculated by

$$St_i = \frac{Cr_i\left(\hat{x}_i - x_i\right)}{\left[\sqrt{\frac{\sum_{j=1}^{n} Cr_j}{n-f}}\right](1 - Leverage_i)} \quad (15)$$

where Cr are the concentration residuals of every sample (actual minus the predicted output value) (i) in the data set. In our data set, we had a (32 x 1) matrix. Therefore 32 leverages and studentized residuals were obtained. The studentized residuals versus the leverages were plotted to determine probable outliers.

The Q-test, another way of finding the outliers can be carried out using the following equation (Forinash, 2007):

$$Q_n = \frac{Y_a - Y_b}{R} \quad (16)$$

where Q_n = at 90% confidence interval Q for n replicate measure, R = range of all data points, Y_a = the suspected outlier, and Y_b= the data point closest to it. In case of our data set, we determined Q value for each individual estimated ET value. Outliers were determined by the highest corresponding Q-value.

A confidence interval for the dataset prediction model ($Y = \alpha + \beta X$) is given by (Bowerman and O'Connel, 1990)

$$C.I. = \hat{Y} \pm t_{\alpha/2} * S_e * \sqrt{\frac{1}{n} + \frac{n(x_i - \bar{x})^2}{S_{xx}}} \quad (17)$$

where $1 - \alpha$ is the confidence interval (C. I.), $t_{\alpha/2}$ is the value calculated from article t-table for n-2 degree of freedom with n as the number of samples, the x_i are the individual data points, *x-bar* is the mean of the data set, and S_e is the standard error of estimate (Bowerman and O'Connel, 1990).

The C.I. test was carried out using the predicted yield and the corresponding seasonal calculated ET data set. The analysis of the test verified the data outside the 95% C.I., showing those as potential outliers. Finally, those outliers were omitted from the yield versus ET data set. A new linear regression model was created.

III. RESULTS AND DISCUSSION

A. *Optimal Corn Yield Prediction Model*

The corn yield optimized RBFN model was obtained with 11-06-0-1-network architecture, i.e., with eleven input neurons, six prototype layer neurons, and one output neuron, corn yield. The same optimal RBFN model was validated by running the model with similar optimal network parameters a minimum of 15 times. The standard deviations of the training and testing RMSEs were 0.02 and 0.10, respectively. The standard deviation of the correlation coefficient between testing actual versus predicted yield was only 0.13 (Table 2). The corresponding standard deviation among predicted and average accuracies of 15 different runs was 1.41. These low standard deviations suggest the stability of the chosen model.

We established a curve-fitting relationship between the actual and predicted crop yield. Average corn yield prediction accuracies of 90.67% and 90.59% were obtained for testing and training datasets, respectively. A correlation coefficient (r) of 0.65 was obtained between actual and predicted yield. The corresponding SEP was 1.13 t/ha (for the test dataset), which corresponds to 9.62% of the average actual yield from the 32 grid plots used in the test dataset. The RMSE of 1.08 t/ha (10.2% of the average actual yield from test grid plots) was obtained from the optimal predicted test model. The maximum and minimum test accuracies obtained from the optimal model were 99.56 and 51.54%, respectively. The minimum, maximum, and the average absolute errors obtained from the test model were 0.04 t/ha, 2.57 t/ha, 0.91 t/ha, respectively. Absolute errors exceeding 1 t/ha in the corn prediction occurred in 11 grid plots. These grid plots were associated with non-irrigated areas in 1998. The RBFN prediction model was useful for predicting the corn yield using imagery and non-imagery data.

Table 3 shows additional performance parameters of the optimal model. The linear regression between actual versus predicted corn yield is shown in Figure 6. The correlation coefficient of 0.65 obtained from this study is similar to the results obtained by other researchers using different corn crop production parameters in different environmental conditions. In this study, the use of four years' data as training datasets for the learning of the network incorporated most of the variations in crop management factors. Gopalapillai and Tian (1999) obtained an average correlation coefficient (r) of 0.54 to 0.79 for predicting corn yield for nine different image bands/indices derived from aerial color infrared (CIR) images in nine different fields. The lowest and highest absolute r-values obtained were 0.11 and 0.99, respectively. They used research test plots with controlled cultivation techniques over two years. Therefore, the corn yield prediction performance using a broad variety of crop production input features (i.e., total environment) shows promise. Another study by Seidel et al. (2000) used the correction normalized vegetation index (NVI) technique to predict soybean yields. The model was developed using only one year (1994) training data from one field, two and three years of training data from other fields, and was tested only with another year. The R^2 (coefficient of determination) between the NVI and the soybean crop yield ranged from 0.05 to 0.69. The lower R^2 obtained with their model may have been caused by the use of the vegetation spectral information only from a single year. Therefore, this study shows the importance of including multiple years to incorporate environmental variation into the corn yield modeling. Even Seidel et al. (2000) obtained good R^2 when they used multiple years input data for soybean yield prediction.

TABLE 2. RESULTS OBTAINED FROM THE OPTIMAL MODEL SEVERAL RUNS AFTER NETWORK INITIALIZATION

Run #	Training		Test		Actual versus predicted yield testing dataset	
	RMSE	Correlation[a]	RMSE	Correlation[a]	Correlation coefficient	Average prediction accuracy (%)
1	0.09	0.97	0.17	0.79	0.58	88.16
2	0.09	0.97	0.17	0.79	0.58	88.16
3	0.09	0.97	0.17	0.49	0.58	88.15
4	0.11	0.96	0.12	0.58	0.65	90.67
5	0.11	0.96	0.12	0.58	0.65	90.68
6	0.06	0.95	0.17	0.25	0.29	87.12
7	0.11	0.96	0.12	0.58	0.65	90.67
8	0.16	0.56	0.13	0.60	0.70	90.68
9	0.09	0.97	0.17	0.28	0.33	87.49
10	0.12	0.66	0.12	0.62	0.66	90.89
11	0.17	0.79	0.12	0.58	0.69	90.68
12	0.12	0.82	0.12	0.62	0.69	90.68
13	0.17	0.61	0.14	0.40	0.46	90.30
14	0.16	0.56	0.13	0.60	0.70	90.68
15	0.12	0.66	0.12	0.62	0.66	90.68
Average	0.12	0.83	0.14	0.56	0.59	89.71
Standard Deviation	0.03	0.17	0.02	0.15	0.13	1.41

[a]Correlation information relates to the actual yield and the model predicted yield at the time of network training and testing.

TABLE 3. SUMMARY OBTAINED FROM NEURAL NETWORK AND REGRESSION ANALYSIS

Model	**Optimal model architecture and parameters (From neural network analysis)**					**Average prediction accuracy (%)**	**Actual-Predicted model correlation (From regression analysis)**				
	Net	LR[a]	M	Iterations	RMSE		α	β	r	SEP (t/ha)	RMSE (t/ha)
Training	11-6-0-1	0.9	0.4	20,000	0.1303	90.59	-3.79	1.36	0.87	1.12	1.11
Test	11-6-0-1	0.9	0.4	20,000	0.1271	90.67	0.38	1.00	0.65	1.02	1.08

[a]Abbreviations: LR = learning rate coefficient, M = momentum coefficient, RMSE = root mean square error, α = intercept, β = slope, r = correlation coefficient, and SEP = standard error of prediction.

B. Water Use Model Development

A linear regression model was created (Figure 7) using these calculated seasonal ET values (mm) versus their corresponding predicted yields (t/ha) obtained from the corn yield test models. The coefficient of determination (R^2) of 0.65 was obtained from the linear regression model. A linear relation may exist between the seasonal cop water use and predicted (forecasted in mid-season) crop yield. However, the correlation coefficient was poor as compared to other studies in this field (Stegman, 1982 and1986; Klocke et al., 1996). Klocke et al. (1996) obtained $R^2$of 0.98 for their actual ET versus actual yield linear model; Stegman (1982 and 1986), had R^2 of 0.70 and 0.90, respectively for his estimated ET versus actual yield linear correlation models. A possible reason for the lower correlation in our study could be attributed to the use of RBFN model predicted crop yield (obtained in mid-season), whereas studies of Stegman and Klocke et al. used the actual crop yields obtained at the end of cropping season. Another possible cause of the lower correlation between ET and predicted yield in our study may be the variability in soil properties within the field. However, we have not used these variations in our study, while estimating the seasonal ET from testing grid plots (32 numbers). Another possible reason for low R^2 was the presence of some outliers in the test dataset.

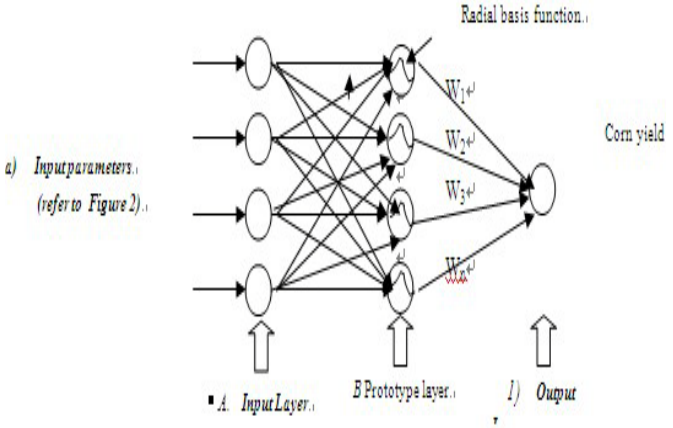

Figure 6. Actual and predicted corn yield correlation curve using the test data.

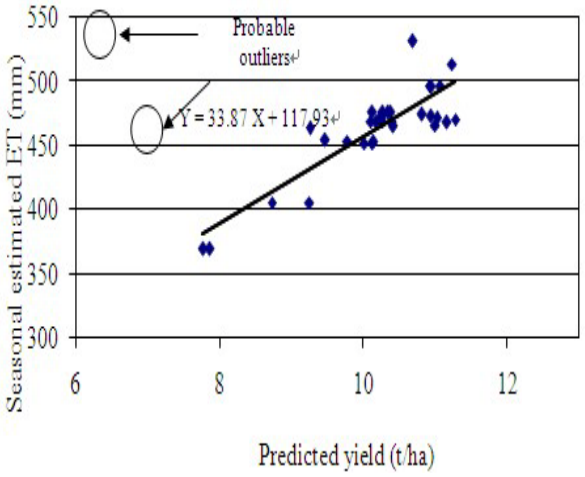

Figure 7.Seasonal estimated ET vs. predicted corn yield (n=32)

The studentized residual test and the Q-test were carried out using the ET data (Figures 8 and 9). The residual analyses suggest that there were four outliers present in the dataset, two from the non-irrigated part of 1997 and one each from the irrigated quarters of 1998 and 2001, respectively. These four data were either exceptionally high or low compared to other 28 ET values. We later confirmed that lysimeters malfunctioned in those plots in the corresponding years. Therefore, we omitted those residuals and another linear ET versus predicted yield regression model was created using the remaining 28 data points, including data from all four years. The revised linear regression curve presented in Figure 10 shows an improved R^2 of 0.81. The 95% confidence intervals were also calculated showing that estimated ET versus predicted yield model fits the data reasonably well.

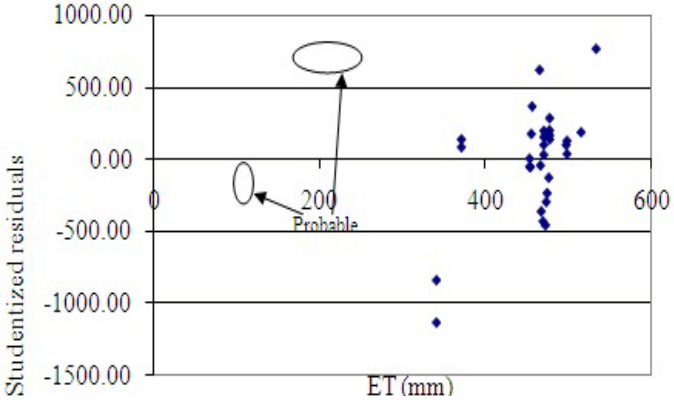

Figure 8.Studentized residual plot to determine outliers in the data set.

Estimated ET was found using the linear fit equation ET = 32.56 x Predicted yield + 129.61 (R^2 = 0.81), where estimated ET has units of mm and predicted yield has units of t/ha. A comparison of actual estimated (found from ET calculation model) and predicted estimated ET (obtained from the ET versus predicted yield regression model) curves along with the prediction errors are provided in Figure 11. The errors in ET prediction for most of the data points were within the range of ±15 mm, with only three data points having an absolute error of more than 20 mm. The maximum absolute error with data points generating predicted yield versus calculated ET was only 26 mm. All higher estimated ET prediction errors were in 1997; predicted yield was higher in that year compared to others. The highest error in prediction of estimated seasonal ET for corn was only 5.58% of the average ET of plots used in the study. Average absolute error in predicted ET using the linear model was 11 mm or 2.4% of average total ET. The minimum error in ET prediction was only 0.94 mm, or 0.2% of average total ET. This study demonstrates that the seasonal water requirement for crops can be predicted using aerial images of crop vegetation, bare soil images, and non-imagery (climatic and non-climatic) data.

Figure 12 shows a comparison of our results with those of Stegman (1982 and 1985) and Klocke et al. (1996). Stegman's (1982) study was conducted in the Oakes, ND area using actual yields and estimated seasonal ET values from 1977, 1978, and 1979. A 1981 through 1983 study in east-central ND (Stegman 1986) produced slope and intercept values of (17.94)

and (282.97), respectively. The R^2 of this study was close to the above-mentioned three studies. The R^2 values from Stegman's 1982 and 1986 studies were 0.71 and 0.90, respectively, compared with R^2 = 0.81 for this study. These results support the validity of the model developed here. The actual yield versus actual ET model developed in Nebraska by Klocke et al. (2002) had a higher R^2 of 0.96, but their model only used three data points. Moreover, the slope (38.80) and intercept (187.14) of Klocke et al.'s (2002) linear model closely matches with the slope (32.56) and intercept (129.6) obtained from the seasonal estimated ET vs. predicted ET linear fit model.

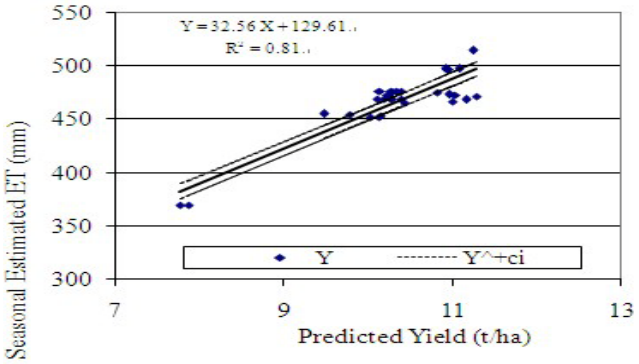

Figure 9.Q-test plot to determine outliers in the data set.

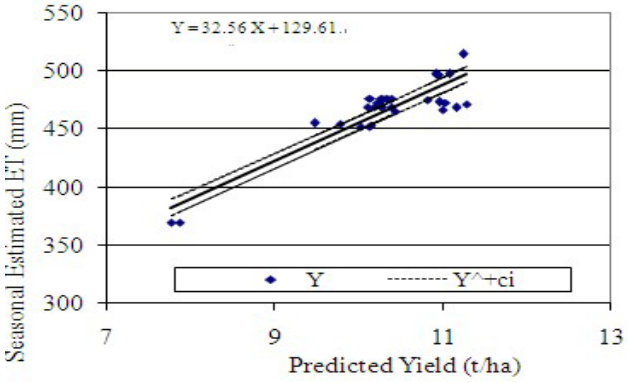

Figure 10. Seasonal estimated ET vs. Predicted corn yield (n=28) after omission of outliers.

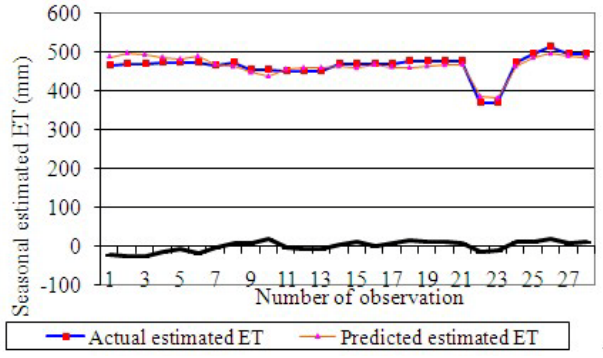

Figure 11. Comparison curve of actual and predicted estimated ET from the ET vs. yield linear regression model (n=28).

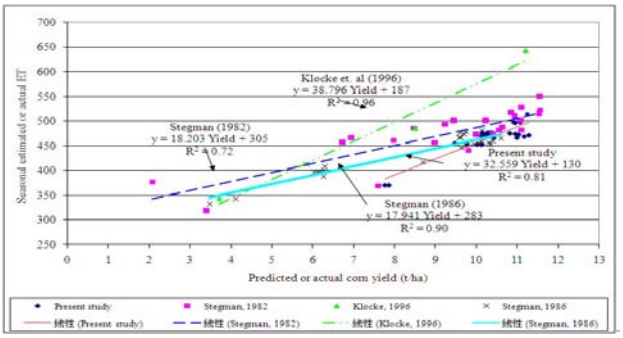

Figure 12.The comparison of the yield versus seasonal estimated ET models with other studies.

IV. SUMMARY AND CONCLUSIONS

Corn yield was estimated with reasonable accuracy using aerial images of crop vegetation, bare soil images, and other climatic and non-climatic data. Principal component band images were used to incorporate variations of individual bands of the visible spectrum (R, G, and B). The RBFN-corn yield prediction model provided an average crop yield prediction accuracy of 91% with actual versus predicted corn yield correlation coefficient (r) of 0.65. Corn yield prediction accuracies could be improved by using other important environmental features, such as wind speed, crop diseases, pest and insect attack information, weed condition in the field, and drought information. Another important crop management factor, applied nitrogen (from basal to the date of image acquisition date) could be used to improve the performance of the corn crop yield model. Features that affect crop production were not considered in this study. We used the mid-season aerial images of corn growing season after another study (Panda 2003) found, after analyzing several images in the growing season, that mid-season images correlated best.

By using the Jensen and Haise equation (1963) and a set of modifying coefficients (Stegman and Coe, 1984) we estimated the seasonal crop water usage from defined spatial grids. There was a reasonable linear relationship between the calculated ET and the estimated crop yield, with R^2 of 0.65. Residual analysis was performed as an effective tool to detect outliers. The studentized residual test and Q-test suggested several probable outliers. After elimination of outliers, the linear relationship between ET and predicted crop yield was improved the R^2 from 0.65 to 0.81.

Data used for developing the seasonal ET model consisted of information from four different years, assuming that this mixing could represent variability of environmental and agricultural production conditions. Additional work needs to be done to determine the capability of the suggested approaches for estimating seasonal ET of a crop in a given year using other information, such as wind speed, wind chill effect, soil water drainage, etc., from the same year/growing season.

REFERENCES

[1] Ali, M.H., Hoque, M.R., Hassan, A.A., Khair, A., 2007. Effects of deficit irrigation on yield, water productivity, and economic returns of wheat. Agricultural Water Management. 92(3), 151-161.

[2] Anonymous. 2002. Optimization: Outlier Sample Detection. http://www.galactic.com/Algorithms/pca_outlier.htm. Accessed on 10.15.07.

[3] Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford, UK.

[4] Bowerman, B.L. O'Connel, R., 1990. Linear Statistical Models. An Applied Approach. PWS-Kent Publishing Company, Boston, MA.

[5] Burrough, P. A., 1993. Soil variability: A late 20th century view. Soils Fertility. 56: 529-562.

[6] Byne, G.F., Crapper, P.F., Mayo, K.K., 1980. Monitoring land cover change by principal component analysis of multi-temporal Landsat data. Remote Sensing of Environment. 10, 175 184.

[7] Canavos, G.C., 1984. Applied Probability and Statistical Methods. Little Brown and Company, Boston, MA.

[8] Derby, N.E., Knighton, R.E., Steele, D.D., 1998. Methods of monitoring leachate losses under irrigated corn best management practices. In: Schaack, J., Freitag, A. W., and Anderson, S. S. (eds). Proceedings of the 1997 Water Management Conference. US Committee on Irrigation and Drainage, Denver, 243-257.

[9] Derenbos, J. Pruitt, W.O., 1977. Guidelines for predicting crop water requirements. Irrigation Drainage Paper No. 24, Food and Agriculture Organization of the United Nations, Rome, Italy, 143.

[10] Forinash, K., 2007. Hints for finding error in your experiment. http://physics.ius.indiana.edu/info/ErrorAnalysis.html. Accessed on 10.15.07.

[11] Fung, T., LeDrew, E., 1987. Application of principal component analysis to change detection. Photogrammetric Engineering and Remote Sensing. 53(12), 1649 1658.

[12] Gallardo, M., Snyder, R.L., Schulbach, K., Jackson, L.E., 1997. Crop growth and water use model for lettuce. Journal of Irrigation and Drainage Engineering, 122(6), 354 – 359.

[13] Goodman, S. D., 1993. A radial basis network for seismic signal discrimination. IEEE Neural Network Conference, 1993, Paris, France.

[14] GopalaPillai, S., Tian, L., 1999. In-field variability detection and yield prediction in corn using digital aerial imaging. Transactions of the ASAE. 42(6), 1911-1920.

[15] Hanks, R.J., 1983. Yield and water use relationships: An overview. In limitations to efficient water use in crop production (H. M. Taylor et al., eds.). American Society of Agronomy. Monograph. Madison, WI, 393-410.

[16] Haykin, S., 1999. Neural networks: A Comprehensive Foundation, Second Edition. Prentice Hall Inc., Upper Saddle River, NJ.

[17] Jensen, M.E., Wright, J.L., Pratt, B.J. 1971. Estimating soil moisture depletion from climatic crop and soil data. Transaction of the ASAE. 14(5), 954-959.

[18] Jensen, M.E., Haise, H.R., 1963. Estimating evapotranspiration from solar radiation. Journal of Irrigation Drainage. 89, 15-41.

[19] Klocke, N.L., Hubbard, K.G., Kranz, W.L., Watts, D.G., 1996. Evapotranspiration (ET) or crop water use. Neb Guide. G90-992-A. www.ianr.unl.edu/pubs/irrigation/g992.htm.

[20] Lee, J., Beach, C.D., Tepedeleliogu, N., 1996. Channel equalization using radial basis function network. IEEE Conference on Signal Processing, 1996.

[21] Lillesand T.M., Keifer, R.W., 1995. Remote Sensing and Image Interpretation. John Willey and Sons, New York, NY.

[22] Moody, J., Darken, C.J., 1989. Fast learning in networks of locally tuned processing units. Neural Computation. 1, 281-294.

[23] Neural Ware, 2000. Reference Guide of NeuralWare. NeuralWare, Carnegie, PA.

[24] Panda, S.S., 2003. Data mining application in production management of crop. Ph.D. Dissertation, North Dakota State University, Fargo, ND, USA.

[25] Ranaweera, D.K., Hubele, N.F., Papalexopoulos, A.D., 1995. Application of radial basis function neural network model for short-term load forecasting. IEEE Proc.-Generation Transmission Distribution. 142(1), 45-50.

[26] Seidel, M.S., Paz, J.O., Batchlor, W.D., 2000. Integrating remotely sensed images to improve spatial crop model calibration. ASAE Paper No. 003057. St. Joseph, MI.

[27] Steele, D.D., Stegman, E.C., and Knighton, R.E., 2000. Irrigation management for corn in the northern great plains, USA. Irrigation Science. 19, 107-114.

[28] Stegman, E.C., 1982. Corn grain yield as influenced by timing of evapotranspiration deficits. Irrigation Science. 3, 75-87.

[29] Stegman, E.C., 1986. Efficient irrigation timing methods for corn production. Transaction of the ASAE. 29(1), 203-210.

[30] Stegman, E.C., Coe, D.A., 1984. Water balance irrigation scheduling based on Jensen_Haise equation: Software for Apple II, II +, and IIE computers. North Dakota Research Report No. 100. ND Agricultural Experiment Station. North Dakota State University, Fargo, ND.

[31] Stegman, E.C., Bauer, A., Zubriski, J. C., Bauder, J., 1977. Crop curves for water balance irrigation scheduling in south eastern North Dakota. North Dakota Research Report No. 66. ND Agricultural Experiment Station. North Dakota State University, Fargo, ND.

[32] Terjung, W.H., Liverman, D.M., Hayes, J.T., O'Rourke, P.A., Todhunter, P.E., 1984. Climatic change and water requirements for grain corn in the North American Great Plains. Climatic Change. 6, 193-220.

[33] Timlin, D. J., Pachepsky, Y., Snyder, V. A., Brynt, R.B., 1998. Spatial and temporal variability of corn grain yield on a hill slope. Journal of Soil Science Society of America. 62,764-773.

[34] Wan, C., Harrington, P.B., 1999. Self-configuring radial basis function neural networks for chemical pattern recognition. Journal of Chemical Information Computer Science. 39, 1049-1056.

[35] Xin, J., Zazueta, F.S., Smajstrla, A.G., Wheaton, T.A., Jones, J.W., Jones, P.H., Dankel II, D.D., 1997. CIMS: An integrated real time computer system for citrus micro irrigation management. Applied Engineering in Agriculture. 13(6), 785-790.

[36] Zhuang, X., Engel, B., 1990. Classification of multi-spectral remote sensing data using neural network vs. statistical methods. ASAE Paper No. 90-7552. St. Joseph, MI